

Open Standards-Based Data Extraction Web Tool for Complex Longitudinal Datasets

Award Information

Agency:

Department of Health and Human Services

Branch

n/a

Amount:

\$150,000.00

Award Year:

2011

Program:

SBIR

Phase:

Phase I

Contract:

1R43AG039898-01

Agency Tracking Number:

R43AG039898

Solicitation Year:

2011

Solicitation Topic Code:

NIA

Solicitation Number:

PA10-050

Small Business Information

ALGENTA TECHNOLOGIES

1428 Washington Ave S, Ste 203, Minneapolis, MN, 55454-

Hubzone Owned:

N

Socially and Economically Disadvantaged:

N

Woman Owned:

N

Duns:

125223888

Principal Investigator:

DAN SMITH

(888) 264-0711

dan@algenta.com

Business Contact:

JEREMY IVERSON

(608) 213-1637

jeremy@algenta.com

Research Institution:

Stub

Abstract

DESCRIPTION (provided by applicant): The NIH funds many long running longitudinal studies that have collected massive amounts of data. These surveys continue to add additional data collection waves to their datasets which increases the wealth of information collected. Unfortunately, as additional waves continue to be added the data becomes more complex for researchers to work with. This is especially true when the study contains many thousands of variables. Researchers are often interested in a subset of the data pertaining to their research question, but have to traverse multiple data files and many pages of documentation to find the variables associated with their topic. It also becomes challenging to replicate research, since it many times involves going through these same burdensome steps. To address this challenge, advanced tools for researchers are needed to navigate and extract data from large longitudinal studies. This project aims to create an open standards based web tool to provide data extracts from large public use longitudinal surveys. The tool will allow researchers to select variables and variable groups to create data extracts. The tool will also create codebook documentation and standardized Data Documentation Initiative (DDI) 3 metadata for the extracts, enabling citation of the extract using the DDI standard. The tool will also be generalized to work for multiple studies by using the DDI open standard for social science research, which is an innovation over today's generation of one-off tools developed on a per study basis. This Phase I feasibility study aims to analyze the data preparation and metadata creation workflow needed to prepare a study for online data extraction, to validate the use of the Data Documentation Initiative's DDI 3 standard for the basis of such a tool, and to create prototype web-based data extraction software. While the focus is on longitudinal surveys, the proposed system would also handle cross-sectional, time-series, and non-repeated studies. The aim is to improve research methodologies through a simplification of the process used for discovering, retrieving, and analyzing data relevant to a researcher's investigation. **PUBLIC HEALTH RELEVANCE:** Researchers who wish to use public use data from longitudinal studies or replicate other's research must currently navigate thousands of variables across multiple waves and datasets to answer simple analysis questions. The proposed web tool allows researchers to create data extracts that are directly related to their queries, allowing more time to be spent on public health research questions instead of data management.

* information listed above is at the time of submission.